Agricultural and Forest Meteorology 206 (2015) 137–150

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape

Aston Chipanshi[a], Yinsuo Zhang[b,*], Louis Kouadio[c,1], Nathaniel Newlands[c], Andrew Davidson[b], Harvey Hill[d], Richard Warren[e], Budong Qian[f], Bahram Daneshfar[b], Frederic Bedard[g], Gordon Reichert[g]

[a] AgroClimate, Geomatics and Earth Observations Division (ACGEO), Science and Technology Branch (STB), Agriculture and Agri-Food Canada (AAFC), 300-2010 12th Ave. Regina, Saskatchewan S4P OM3, Canada
[b] ACGEO, STB, AAFC, 960 Carling Avenue, Ottawa, Ontario K1A 0C6, Canada
[c] Lethbridge Research Centre, STB, AAFC, 5403 1st. Ave. S., Lethbridge, Alberta T1J 4B1, Canada
[d] ACGEO, STB, AAFC, 107 Science Place, Saskatoon, Saskatchewan S7N 0X2, Canada
[e] ACGEO, STB, AAFC, 850 Lincoln Road, Fredericton, New Brunswick E3B 4Z7, Canada
[f] Eastern Cereal and Oilseed Research Centre, STB, AAFC, 960 Carling Avenue, Ottawa, Ontario K1A 0C6, Canada
[g] Remote Sensing and Geospatial Analysis, Agriculture Division, Statistics Canada, Ottawa, Ontario K1A 0T6, Canada

ARTICLE INFO

Article history:
Received 20 June 2014
Received in revised form 9 March 2015
Accepted 14 March 2015
Available online 26 March 2015

Keywords:
Probabilistic yield forecast
ICCYF
Canada
Spring wheat
Barley
Canola

ABSTRACT

Early warning information on crop yield and production are very crucial for both farmers and decision-makers. In this study, we assess the skill and the reliability of the Integrated Canadian Crop Yield Forecaster (ICCYF), a regional crop yield forecasting tool, at different temporal (i.e. 1–3 months before harvest) and spatial (i.e. census agricultural region – CAR, provincial and national) scales across Canada. A distinct feature of the ICCYF is that it generates in-season yield forecasts well before the end of the growing season and provides a probability distribution of the forecasted yields. The ICCYF integrates climate, remote sensing derived vegetation indices, soil and crop information through a physical process-based soil water budget model and statistical algorithms. The model was evaluated against yield survey data of spring wheat, barley and canola during the 1987–2012 period. Our results showed that the ICCYF performance exhibited a strong spatial pattern at both CAR and provincial scales. Model performance was better from regions with a good coverage of climate stations and a high percentage of cropped area. On average, the model coefficient of determination at CAR level was 66%, 51% and 67%, for spring wheat, barley and canola, respectively. Skilful forecasts (i.e. model efficiency index > 0) were achieved in 70% of the CARs for spring wheat and canola, and 43% for barley (low values observed in CAR with small harvested area). At the provincial scale, the mean absolute percentage errors (MAPE) of the September forecasts ranged from 7% to 16%, 7% to14%, and 6% to 14% for spring wheat, barley and canola, respectively. For forecasts at the national scale, MAPE values (i.e. 8%, 5% and 9% for the three respective crops) were considerably smaller than the corresponding historical coefficients of variation (i.e. 17%, 10% and 17% for the three crops). Overall, the ICCYF performed better for spring wheat than for canola and barley at all the three spatial scales. Skilful forecasts were achieved by mid-August, giving a lead time of about 1 month before harvest and about 3–4 months before the final release of official survey results. As such, the ICCYF could be used as a complementary tool for the traditional survey method, especially in areas where it is not practical to conduct such surveys.

1. Introduction

Driven by the increasing societal awareness of the impacts of extreme weather events on crop yields and quality, and the increasing information demand from producers, grain traders, transporters and government policy makers regarding market access and food security planning, many countries have

* Corresponding author. Tel.: +1 6137155026.
E-mail address: yinsuo.zhang@agr.gc.ca (Y. Zhang).

[1] Present address: International Centre for Applied Climate Sciences, University of Southern Queensland, West Street, Toowoomba, 4350, Australia.

http://dx.doi.org/10.1016/j.agrformet.2015.03.007

developed crop monitoring and yield forecasting systems to provide regional, national and global outlooks. Traditionally, regional or national crop yield estimates are made by field surveys and/or farmer interviews conducted during or after the crop growing season (USDA, 2012; Statistics Canada, 2014). Although the survey method is employed in many operational crop yield and production reporting systems, it has several challenges which include: (1) restrictions in resources to meet frequent sampling times and sample size required to get a reliable survey; (2) demands to increase the lead time of the survey reports; (3) needs to lower the questionnaire burden for survey respondents and (4) reliability concerns associated with sampling and non-sampling errors in data gathering and data processing (Statistics Canada, 2014).

To address some of the problems associated with the survey method, tremendous efforts have been made to provide crop yield forecasts from Earth Observation (EO) datasources (Nikolova et al., 2012; Basso et al., 2013; van Ittersum et al., 2013; Johnson, 2014). Normally, mathematical models are constructed to build relationships between crop yields and EO data based predictors. Mathematical models however, face difficulties too in terms of finding the correct combination of predictors at spatial scales that are relevant from an economic point of view. In a comprehensive review of crop yield forecasting methods, Basso et al. (2013) classified mathematical forecasting models into three categories, i.e. statistical (e.g. Thompson, 1969; Qian et al., 2009b; Bornn and Zidek, 2012), mechanistic (e.g. Moulin and Beckie, 1993; Chipanshi et al., 1999; Folberth et al., 2012) and functional types (e.g. Supit, 1997; Ritchie and Alagarswamy, 2002; van Diepen et al., 2004; Basso et al., 2012). Statistical models relate crop yield to various environmental variables by simple regressions (e.g. Thompson, 1969; Mkhabela et al., 2011) or more sophisticated non-linear statistical algorithms (e.g. Bornn and Zidek, 2012; Johnson, 2014). By the statistical estimation of unknown predictor variables, statistical models become amenable to providing within-season probabilistic yield projections well before harvest (Potgieter et al., 2003). Statistical models normally require identifying variables that contribute the most variance to the final yield (e.g. Qian et al., 2009b; Johnson, 2014). Mechanistic models use biophysical principles of plant, soil, weather and management interactions to simulate the crop yield. Because of the need to describe biophysical processes using the best science available, there is a considerable requirement to have detailed crop, soil and management parameters to allow the processes to be expressed in mathematical terms. Often times, the data is sparse and not available at the regional level. Functional models are simplified versions of the complex mechanistic models or a combination with statistical schemes and are more suitable for operational crop yield forecasting due to their minimal data input requirements and the key processes can be parametrized using approximate equations (Basso et al., 2013).

Most EO based methods utilize predictors from either climate derived indices (e.g. Campbell et al., 1997; Supit, 1997; Potgieter et al., 2003; Qian et al., 2009b; Bornn and Zidek, 2012; Szulczewski et al., 2012), or remote sensing derived index, such as the normalized difference vegetation indices (NDVI) (e.g. Bullock, 1992; Boken and Shaykewich, 2002; Doraiswamy et al., 2004; Mkhabela et al., 2011; Johnson, 2014). In recent years, functional models have tended to integrate both climate and remote sensing indices into unified yield prediction models (Prasad et al., 2006; De Wit, 2007; Schut et al., 2009; Cammarano, 2009). Despite the long history of the EO based yield forecasting methodologies, it is only in the last decade or so, when EO datasets became available in near real time (NRT) from open sources (e.g. Han et al., 2012; Joint Research Centre, 2012; NASA, 2013; Atzberger, 2013), that the EO based forecasting methods have become practical enough to assist with the in-season crop yield prediction. Examples of EO based NRT crop monitoring and yield reporting systems include (1) the Crop Condition Assessment Program (CCAP) of Statistics Canada (Reichert and Caissy, 2002), (2) Monitoring Agriculture with Remote Sensing (MARS) Crop Yield Forecasting System (MCYFS) by the European Commission Joint Research Centre (JRC) (van Diepen et al., 2004), (3) China Crop Watch (Wu et al., 2014), (4) the state and shire commodity forecasts by the Queensland Alliance for Agriculture and Food Innovation (QAAFI) and the Department of Agriculture and Food of Western Australia (DAFWA) (Nikolova et al., 2012), (5) the yield forecasting program of the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) (Han et al., 2012; Johnson, 2014) and (6) the Integrated Canadian Crop Yield Forecaster (ICCYF) of Agriculture and Agri-Food Canada (Chipanshi et al., 2012). Some of these efforts are now integrated into the global early warning systems to assess food security around the globe (USAID, 2014; FAO, 2014).

Compared to the traditional survey method, EO data based crop yield forecasting models have several advantages: the human related biases or errors are reduced as most EO data types are obtained by automated instruments; the spatial and temporal coverage are improved as the EO data are obtained more frequently (e.g. daily climate and weekly NDVI) and they can cover the entire crop land rather than a fraction of an area; the lead time for in-season yield forecasts has increased as the NRT data for yield models can normally be available in about a week from their acquisition date; and the EO methods are cost effective as more and more EO based data become publicly available (e.g. Joint Research Centre, 2012; NASA, 2013; Statistics Canada, 2013b; Environment Canada, 2014).

In spite of the improvements in data acquisition, many significant challenges still exist with EO data based functional models. Firstly, effective integration of different EO datasets into one modelling platform remains a challenging task (Nikolova et al., 2012; Basso et al., 2013; Johnson, 2014); secondly, most of the existing yield forecasting models are limited to a few crops or certain geographic regions (Nikolova et al., 2012; Huang and Han, 2014); thirdly,the forecast model's limitation and uncertainty need to be quantified by crop type, forecasting lead times, different geographic regions and other factors (e.g. climate, soil and management practices etc.), before the model can be used for real world operational applications (Nikolova et al., 2012; Basso et al., 2013; Huang and Han 2014) and fourthly, the scaling up/down of yield forecasts may be complicated by the data availability at the scales involved and the choice of a suitable technique (Nikolova et al., 2012; Basso et al., 2013).

An attempt was made to address some of the above challenges in the ICCYF tool. The ICCYF employs a simple process-based versatile soil moisture budget (VSMB) model to assimilate climate data (air temperature and precipitation) with soil and crop variables (Baier et al., 2000). The VSMB derived variables, such as growing degree days (GDD) and soil water deficit indices are further integrated with NDVI datasets from satellite platforms using statistical algorithms. Bayesian statistics and Markov-chain Monte Carlo (MCMC) algorithms are used to generate in-season probabilistic forecasts well ahead of the harvest time (Newlands and Zamar, 2012). The description of the modelling methodology of the ICCYF and its application for spring wheat yield forecasting on the Canadian Prairies were presented in Newlands et al. (2014). The objectives of this paper thus, are (1) to further evaluate the ICCYF model in forecasting the yields of three major grain crops (spring wheat, barley and canola) across the entire Canadian agricultural landscape and (2) to test two aggregating methods in achieving the provincial and national scale yields from the basic modelling units, i.e. the census agriculture region (CAR) used in Statistics Canada's census of agriculture data collection and dissemination activities (Statistics Canada, 2012). The three test crops were chosen on the

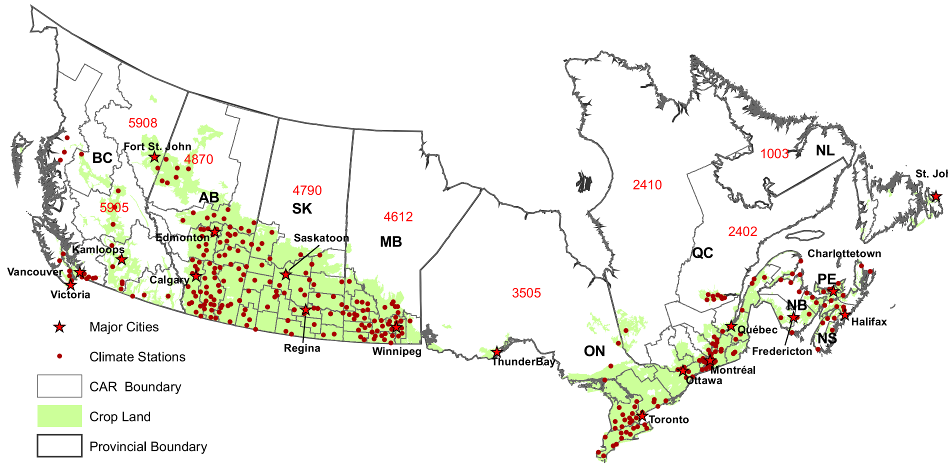


Fig. 1. Census agricultural regions (CARs), crop land extent and selected climate stations across Canadian agricultural landscapes. BC, AB, SK, MB, ON, QC, NB, PE, NS and NL refer to the province of British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Quebec, New Brunswick, Prince Edward Island, Nova Scotia and Newfoundland and Labrador, respectively. The 4 digit numbers indicate the CAR unit identifier.

basis of their economic significance in Canada and the similarity of crop parameters in the VSMB model simulations.

2. Materials and methods

2.1. Study region

This study covers the entire agricultural landscapes of Canada, which encompasses a vast area from 42°N to 60°N in latitude and from 60°W to 130°W in 1ongitude (Fig. 1). Because of the large geographical extent, Canada's agricultural landscape has a wide range of climate types (Phillips, 1990). The climate of British Columbia (BC) is Mediterranean along the Pacific coast and is more continental in the interior dry land. The Prairies provinces, Alberta (AB), Manitoba (MB) and Saskatchewan (SK), have a typical continental climate – very cold winters, hot summers and relatively low precipitation. Southern Ontario (ON) and Quebec (QC) have a climate with hot, humid summers and cold, snowy winters. Provinces along the Atlantic coast, i.e. New Brunswick (NB), Prince Edward Island (PE) and Nova Scotia (NS), have amaritime climate, characterized by high precipitation and mild temperatures. The average annual precipitation ranges from under 400 mm in some part of the Prairies to more than 2400 mm along the BC coast (Phillips, 1990). The estimated average growing season length ranges from 97 to 156 days and the growing season heat accumulation ranges from 647 to1700 degree days (base temperature 5 °C) for cool season crops (Qian et al., 2013).

There are ten soil orders classified by the Canadian System of Soil Classification across the landmass of Canada (Canadian System of Soil Classification Working Group, 1998) and only a few of them dominate the agricultural lands of Canada: Chernozemic soils in the grassland regions of Canada including the great expanse of the Canadian prairies, Luvisolic soils in Southern Ontario, Gleysolic soils in South-Western Ontario, and the Brunisolic soils in the St. Lawrence valley and Southern Manitoba. Due to large variations in soil properties, the soil properties related to crop growth can vary dramatically from place to place. For example, the root zone soil available water holding capacity (AWHC) varies from 50 mm to 250 mm across the agricultural soil zones as estimated from the database of soil landscapes of Canada (Shields et al., 1991).

The spatial scales in this study were determined by census and administrative boundaries, i.e. Statistics Canada's CAR, provincial and national scales, as they are the scales at which most socio-economic data and crop statistics are available to decision-makers. There were a total of 82 CARs in the 2011 census of agriculture across the entire agricultural landscapes of Canada (Fig. 1). However, for less extensive crops, Statistics Canada reports yield by aggregated regions (e.g. most crops in CARs 5901–5907 were reported as one combined unit) or reported only at provincial unit. This is particularly true for spring wheat and canola in Eastern Canada and the Maritime provinces where farm units are smaller than the Prairie provinces. Table 1 lists the yield report units and harvested area of the three crops in this study. The three Prairie provinces (AB, SK and MB) account for 99%, 92% and 99% of the national total harvested area for spring wheat, barley and canola, respectively.

2.2. Input data and processing methods

The primary data inputs for model calibration and validation are the crop survey data reported by Statistics Canada and the EO data (NDVI and climate) from 1987 to 2012. Although both the climate and crop yield data had a longer time series, the study period was chosen according to the availability of the NDVI data series. The data processing procedures are illustrated in Fig. 2 (upper box). Historical crop yield and harvest area for each of the reporting units were obtained through a sample survey of small crop area (i.e. farm level) with a cross-sectional design (Statistics Canada, 2013a). The survey data are weighted and two level indicators are established for the production and harvested area at the CAR and provincial scales. The average yield at those two scales is then calculated. The methodology and error control of the crop survey are explained in the online documents accompanying the Field Crop Reporting Series (Statistics Canada, 2014). Yield and harvest area data are sometimes unpublished due to the unacceptable large coefficient of variation (e.g. greater than 25% CV) or data suppression in order to comply with the privacy laws (Statistics Canada, 2014).

NDVI values were derived from the National Oceanographic and Atmospheric Administration's (NOAA) Advanced Very High Resolution Radiometer (AVHRR) platform (spatial resolution of 1 km in the agricultural regions of Canada) (NOAA, 2013). Cloud removal and other quality control measures were applied to the raw data (Reichert and Caissy, 2002). A crop land extent map, which was derived from the land cover for agricultural regions of Canada (Agriculture and Agri-Food Canada, 2005), and the CAR boundary map (Statistics Canada, 2012) were used to mask the pixel based remote sensing NDVI and extract the CAR level mean NDVI values. More details on the weekly NDVI composite process can be found

Table 1
Number of crop reporting units and the percentage of harvested area over the national total harvested area (PHAN) for each province and the Prairies (Manitoba, Saskatchewan and Alberta).

Province	Spring wheat		Barley		Canola	
	Reporting units[a]	PHAN[b] (%)	Reporting units	PHAN (%)	Reporting units	PHAN (%)
Prince Edward island (PE)	1	0.1	1	0.8	0	0
Nova Scotia (NS)	1	0	1	0.1	0	0
New Brunswick (NB)	1	0	1	0.3	1[c]	0.1[c]
Québec (QC)	1	0.5	11	3.2	1	0.2
Ontario (ON)	1	0.5	5	3.2	1	0.5
Manitoba (MB)	12	17.1	12	10.7	12	18.4
Saskatchewan (SK)	20	53.6	20	37.2	20	47.3
Alberta (AB)	8	27.9	8	43.7	8	33.1
British Columbia (BC)	2	0.4	2	0.8	2	0.7
The Prairies (MB,SK,AB)	40	98.6	40	91.6	40	98.8
Canada	47	100	61	100	45	100

[a] Number of reporting units are extracted from the crop small area data by Statistics Canada (2013a).
[b] Calculation is based on the mean harvested areas over all the available years during 1987–2012, which are extracted from CANSIM, table 001-0010 (Statistics Canada, 2013b).
[c] Canola yield and harvest area reporting in NB has started since 2012, thus, excluded from this study.

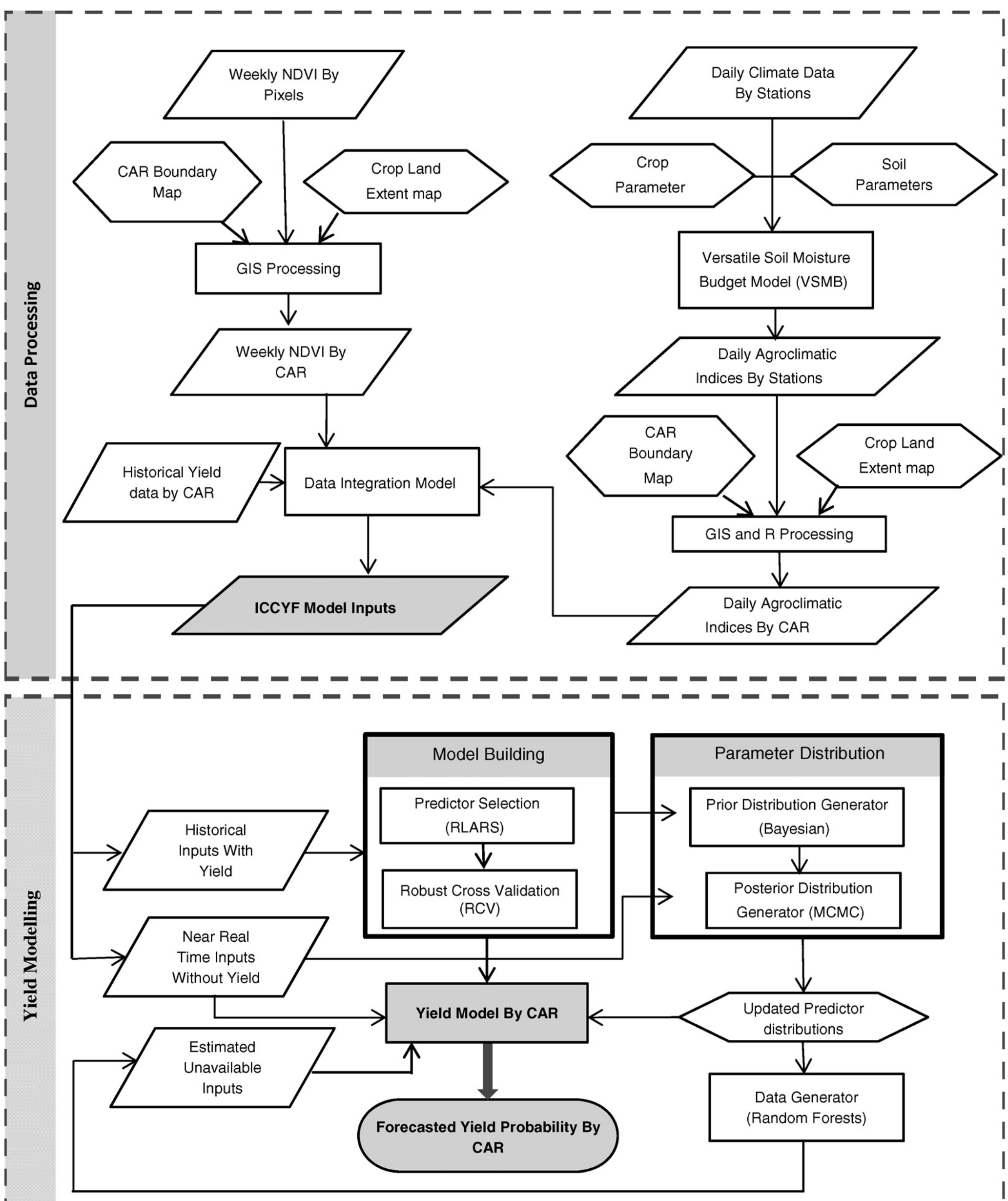


Fig. 2. Data and model flowchart of the Integrated Canadian Crop Yield Forecaster (ICCYF). CAR: census agricultural region; GIS: geographic information system; RLARS: robust least angle regression scheme; MCMC: Monte Carlo Markov-chain.

in Statistics Canada (2010). The weekly AVHRR NDVI data used in this study spanned the 1987–2012 period and covered the growing season from Julian week 18–40 (from the beginning of May to the end of September).

The station based daily temperature and precipitation data provided by Environment Canada and other partner institutions were used to generate the climate based predictors. In total, 330 climate stations across the crop land extent of Canada were selected to represent the climate of the 82 CARs (Fig. 1). The quality control and gap-filling of the missing data was achieved by the in house NRT data quality assurance and control methods at AAFC. In the ICCYF, daily series of air temperature and precipitation from 1987 to 2012 for the 330 stations were input into the VSMB model (Baier et al., 2000) to generate the agroclimatic indices used in the yield forecasting model. The soil physical parameters (e.g. AWHC) required by VSMB were obtained from the Canadian Soil Information Service (CanSIS) database (Shields et al., 1991). A Bio-Meteorological Time (BMT) scale was taken from Robertson (1968) to quantify the crop phenological development. Spring wheat crop parameters for BMT modelling and water extraction coefficients as in Akinremi et al. (1996) were used to produce generalised agroclimatic indices for spring wheat, barley and canola. The outputs generated at a daily time step by the VSMB model and used as potential yield predictors are: Growing Degree Days (GDD) above a base temperature of 5 °C, Soil Water Availability (SWA) expressed as the percent of AWHC, a crop Water Deficit Index (WDI) defined as WDI = 1 – AET/PET, where AET and PET are the simulated actual and potential evapotranspiration respectively and the Crop Seeding Date (CSD) estimated from heat accumulation and top layer soil moisture condition in spring (Baier et al., 2000). Precipitation (P) is also included as a potential yield predictor. Average values of the indices at all stations within the cropland of a specific CAR are used to represent the mean agroclimate of that CAR. If a CAR lacks input climate data, stations from neighbouring CARs are used.

In order to be used in the statistical yield forecasting model, the daily agroclimatic indices were further aggregated into monthly sums (GDD and P) and means (SWA and WDI). The standard deviations (Std) of daily temperature, precipitation and WDI over each month were also calculated and included as potential predictors. The larger the Std value, the higher the variability of the parameter in that month. The weekly NDVI were aggregated into 3-week running means to form the input matrix with the historical yield and the aggregated agroclimatic indices. The monthly time frame for agroclimate aggregation corresponds to the current yield forecast report release time intervals. The 3-week time frame for NDVI data aggregation was based on the sensitivity studies by Hochheim and Barber (1998) and Mkhabela et al. (2011), which showed that the 3-week time integration of NDVI provided an adequate balance between yield sensitivity and model stability.

2.3. ICCYF and its statistical algorithms

Four major steps are followed when generating the in-season yield forecasts (Fig. 2, bottom box): (1) building a statistical prediction model at the CAR level, (2) generating the statistical distribution of parameters, (3) estimating unknown values of predictor variables and (4) forecasting the probabilistic crop yield. To build a customised model at the CAR level, the historical annual series of crop yield were treated as yield observations while the annual series of 3-week NDVI values and monthly agroclimatic indices were treated as potential predicators. All predictors were put into a robust least angle regression scheme (RLARS) (Efron et al., 2004; Khan et al., 2007) to evaluate and rank those variables that contribute the majority of the variance in the predicted yield. A maximum number of predictors (currently set at five) was set based on the sample size of the model building data

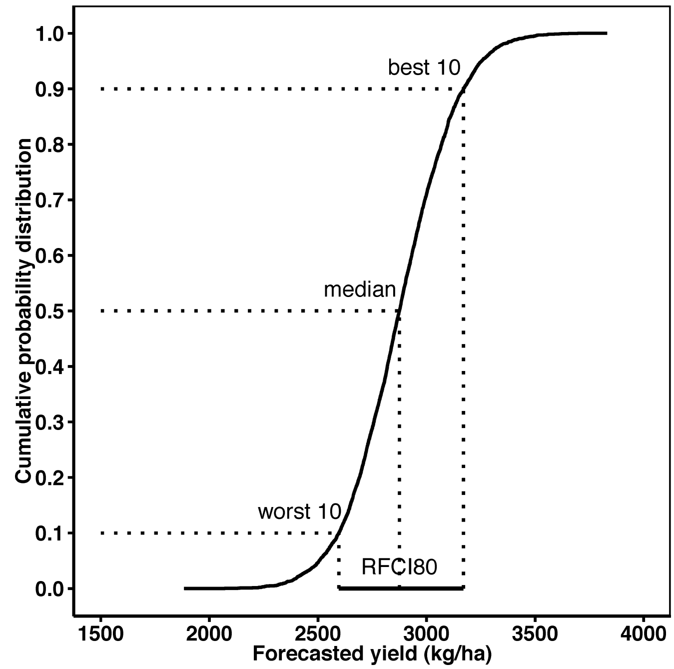


Fig. 3. Example of forecasted yield distribution obtained through the Integrated Canadian Crop Forecaster for the census agricultural region 4791. Dotted lines indicate the positions of forecasted 10th percentile (worst 10), 50th percentile (median) and 90th percentile (best 10) of the cumulative yield distribution curve. The horizontal thick black line between the worst 10 and best 10 lines represents the range of forecasted 80% confidence interval (RFCI80).

(Newlands et al., 2014). The selected predictors were then subjected to a robust cross validation (RCV) scheme (Khan et al., 2010) to finalise the predictors and coefficients of each yield model. The RCV process further stabilized the model by eliminating any false predictors selected from contaminated data (Khan et al., 2010). Based on Bornn and Zidek (2012), incorporating spatial correlation among the yield models of neighbouring CARs increases the individual model's prediction power and stabilizes the model performance. Therefore, we adopted the Bayesian statistical approach as described by Bornn and Zidek (2012) for spatial correlation analysis. Historical data of both forecasting CAR and statistically selected neighbouring CARs were used to establish the prior distribution of the predictors. The posterior distribution of the predictors was obtained using the Markov-chain Monte Carlo (MCMC) scheme (Dowd, 2006). The unobserved variables that are required to make a forecast by a specified date within the growing season were estimated from a statistical procedure called Random Forest Algorithm (Liaw and Wiener, 2002). The estimated variables and the variables observed at near real time were then used as input into the selected yield model to forecast the yield probability distribution for each CAR. An example of the forecasted yield probability distribution is illustrated in Fig. 3. Similar to the probability measures used by Potgieter et al. (2003), the 10th percentile (worst 10%), the 50th percentile (median) and the 90th percentile (best 10%) were output as the probability measures in this study. The range of the 80% forecast confidence interval (RFCI 80) was represented by the distance between the forecasted best 10% yield and the worst 10% yield.

The basic prediction model at the CAR level is a multivariate linear equation defined as:

$$Y_t = \alpha_0 + \alpha_1 t + \sum_{i=2}^{n} \alpha_i X_{i,t} + \epsilon_t \quad (1)$$

where Y_t is the crop yield of year t, α_0 is the regression intercept, $\alpha_1 t$ represent the technology trend of yield over time, $X_{i,t}$ is the predictor i in year t, i could be any of the potential predictors, such as NDVI or agroclimatic indices in any of the averaging periods, ϵ_t is the error term. Eq. (1) is a simplified version of Eq. (11) in Newlands et al. (2014), in which, the yield of previous year (Y_{t-1}) was also considered as a potential predictor. However, Y_{t-1} was only selected occasionally by very few CARs during model validation, thus, this variable was excluded from further analysis in this study.

2.4. Yield forecasting at provincial and national scales

For provinces with only one reporting unit (e.g. PE, NS and NB, Table 1), the provincial yield is the unit of analysis. For those with multiple reporting units (MB, SK, AB and BC and QC and ON for barley, Table 1), the provincial yields were forecasted with the following two methods:

(1) Yield-aggregating method (YAM): The yield at each individual reporting unit (Y_j) is modelled separately and the mean yield at the provincial scale ($\bar{Y}$) is achieved by weight-averaging the yields of all the reporting units (m) of that province using harvested area as a weighting factor:

$$\bar{Y} = \sum_{j=1}^{m} Y_j S_j / \sum_{j=1}^{m} S_j \quad (2)$$

- Predictor-aggregating method (PAM): All the predictors ($X_{i,j}$) are firstly averaged to provincial scale ($\bar{X}_\iota$) using harvest area as weighting factor (Eq. (3)). Where i = 1, n are all the potential predictor as in Eq. (1) and j = 1, m are all the yield reporting units of the province.

$$\bar{X}_\iota = \sum_{j=1}^{m} X_{i,j} S_j / \sum_{j=1}^{m} S_j \quad (3)$$

All the historical $\bar{X}_\iota$ and historical yields at provincial scale are then used to build yield models at provincial scale using Eq. (1) and the provincial yields are then directly forecasted at the provincial scale.

Both YAM and PAM methods were used in an attempt to find an approach that best forecasts the yields at the provincial and national levels.

2.5. Model validation

Existing measures in assessment of forecasts are quite diverse due to the differences in perception of a 'good' forecast between the forecast users and the forecasters and even among individuals of any of the two groups (Murphy, 1993; Potgieter et al., 2003; Krause et al., 2005; Szulczewski et al., 2012). Since the assessment of forecast value involves considerable user inputs and is beyond the scope of this study, we focused on measures to assess the forecast quality and consistency of the ICCYF in forecasting the yields of spring wheat, barley and canola. However, the quality and consistency are multifaceted in nature and numerous statistical indices are used to address their measures (Potgieter et al., 2003; Krause et al., 2005; Szulczewski et al., 2012). In addressing quality and consistence measures, we tried to answers the following questions: (1) how credible are the forecasts associated with different crops, lead times and crop regions? (2) Are the forecasts better than common judgements (e.g. historical mean)?

Among the many quality and consistency measures used in various model evaluation studies (e.g. Potgieter et al., 2003; Krause et al., 2005; Szulczewski et al., 2012), we calculated and compared those indices that are relevant to the questions aforementioned, that is, Bravais and Pearson coefficient of determination (R^2), Root mean squared error (RMSE), relative RMSE (RRMSE), coefficient of residual mass (CRM), mean absolute percentage error (MAPE) and model efficiency index (MEI). Initial results showed that some of these indices were highly correlated (e.g. MEI and CRM, RRMSE and MAPE) and thus, only one from each group was chosen. Finally, two quality measures (R^2 and MAPE) and one consistency measure (MEI) were selected to validate results from the ICCYF.

R^2 is defined as:

$$R^2 = \left(\left(\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P}) \right) / \left(\sqrt{\sum_{i=1}^{n} (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^{n} (P_i - \bar{P})^2} \right) \right)^2 \quad (4)$$

where O_i and P_i are observed and forecasted yield at forecast year i, a bar represents a mean value. R^2 is one of the most frequently used model performance indicators. Its value lies between 0 and 1, which describes how much of the observed variance is explained by the predictors. However, R^2 is sensitive to sample size (n) and extreme values and thus, require some hypothesis testing (e.g. t-test) to better understand its statistical significance. In spite of these problems, R^2 is still one of the most frequently used model evaluation measures (e.g. Boken and Shaykewich, 2002; Mkhabela et al., 2011; Johnson, 2014), especially during the model calibration phase.

MAPE is an accuracy measure of the forecast quality. It is calculated as:

$$\mathrm{MAPE} = 100 \times \frac{1}{n} \times \sum_{i=1}^{n} \left(|O_i - P_i| / |O_i| \right) \quad (5)$$

As a percentage (relative) error measure, MAPE is a better index than absolute error measures (e.g. RMSE) in comparing model performance among different regions and crops, as their historical means could differ considerably.

The MEI, or Nash–Sutcliffe efficiency index (Krause et al., 2005; Szulczewski et al., 2012) is defined as:

$$\mathrm{MEI} = 1 - \sum_{i=1}^{n} (O_i - P_i)^2 / \sum_{i=1}^{n} (O_i - \bar{O})^2 \quad (6)$$

MEI uses variance of the observed values to normalize the forecast errors and its value can range from $-\infty$ to 1. An efficiency index of 1 corresponds to a perfect match of forecasted and observed yield. An efficiency index of 0 indicates that the model predictions are as good as the average, whereas, MEI less than zero means that the forecast is inferior to the use of the historical average yield as a forecast. MEI could be considered as a consistency measure or a measure of model skill as defined by Murphy (1993). Essentially, the closer the MEI is to 1, the more skilful the model is. When MEI < 0, the model should be recalibrated or replaced with a historical mean.

Evaluation measures obtained during the model calibration usually give some indication of how the yield correlated to the variables selected in the model. However, due to over-fitting or biases in the input data, a skilful model at calibration might not give good forecast results once new input data sets are introduced (Qian et al., 2009b; Schut et al., 2009). An improved evaluation approach is to set aside a sub-set of data for validation and use the rest (training data) to calibrate the model. If sufficient data is available for both training and validation, this approach will provide reliable model evaluations as the training data and model building/calibration data are independent of each other. This approach is however limited by sample size. For example, in this study, NDVI were available for 26 years (1987–2012) and insufficient for both training and validation. Cross validation, particularly leave-one-out-cross-validation (LOOCV), has been proven to be effective for model evaluations with limited sample size (Efron, 1983; Khan et al., 2010), and has been used in similar studies (e.g. Qian et al., 2009b; Schut et al., 2009; Mkhabela et al., 2011). The LOOCV method was adopted in this study to assess the performance of the ICCYF for the three crops at three spatial scales. R^2, MEI and MAPE were compared for all test

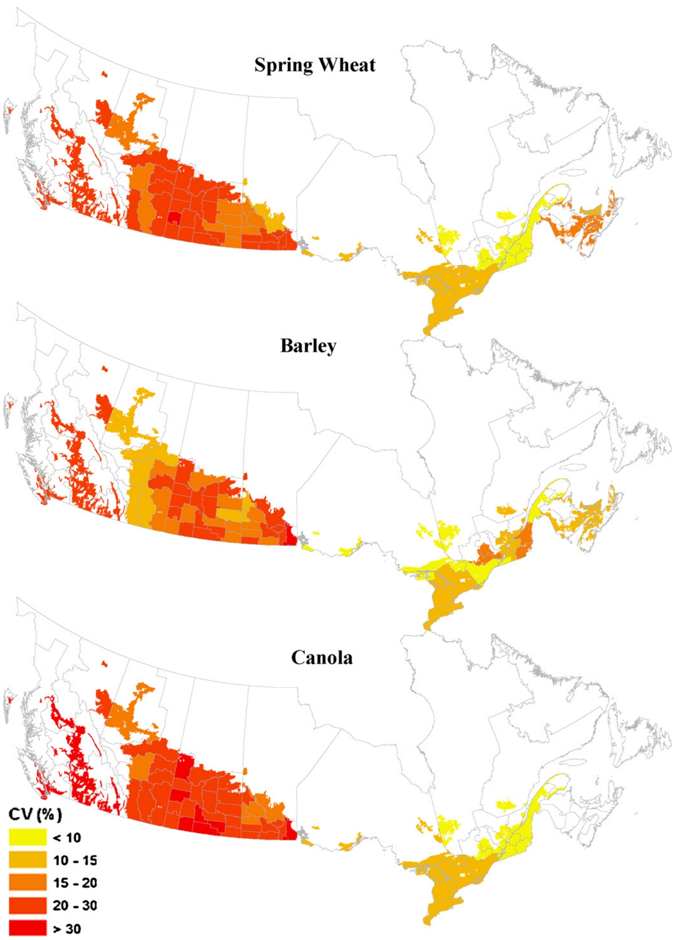


Fig. 4. Coefficient of variation (CV) of the surveyed yield of spring wheat, barley and canola over 1985–2012 period at the census agricultural region (CAR) scale across Canada. Only the crop land within a given CAR was used to map the results.

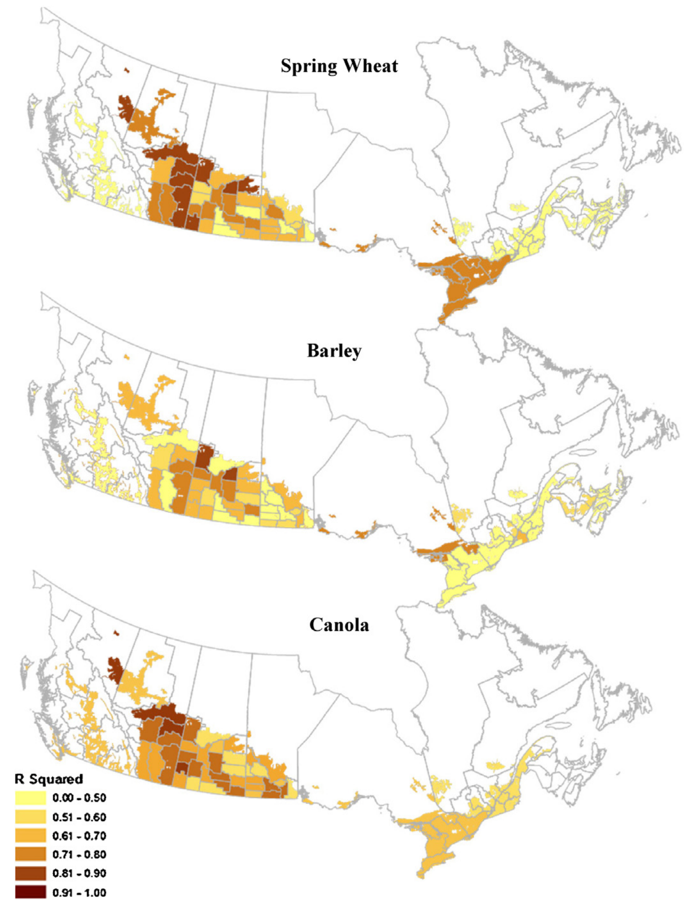


Fig. 5. Distribution of Bravais and Pearson coefficient of determination (R^2) during model calibration at census agricultural region (CAR) level across Canada for the three tested crops. Only the crop land within a given CAR was used to map the results.

runs at all scales. Model predictors being selected for the three crops at all the regions were also analysed.

3. Results

3.1. Observed yield variation at CAR level

The variation of yield in historical records is important for assessing the yield forecast results because regions with highest yield variation are challenging in yield forecasting efforts. The coefficient of variation (CV), which is expressed as the ratio of the standard deviation over the mean, is used to quantify the observed yield variation over the study period (1987–2012). The CV distributions of spring wheat, barley and canola across Canada's agricultural landscape (Fig. 4) show that the yields for all three crops were more variable in Western Canada (from Manitoba to British Columbia) than in Eastern Canada (Ontario and Eastward to Atlantic Canada). Most Prairie CARs exhibited CVs of more than 20%, implying the higher dependence of yield on environmental conditions. The variation in historical yield of canola was the highest, followed by those for spring wheat and barley. The number of CARs with a CV exceeding 20% was 36, 29 and 21, for the three respective crops.

3.2. Yield correlation at CAR level

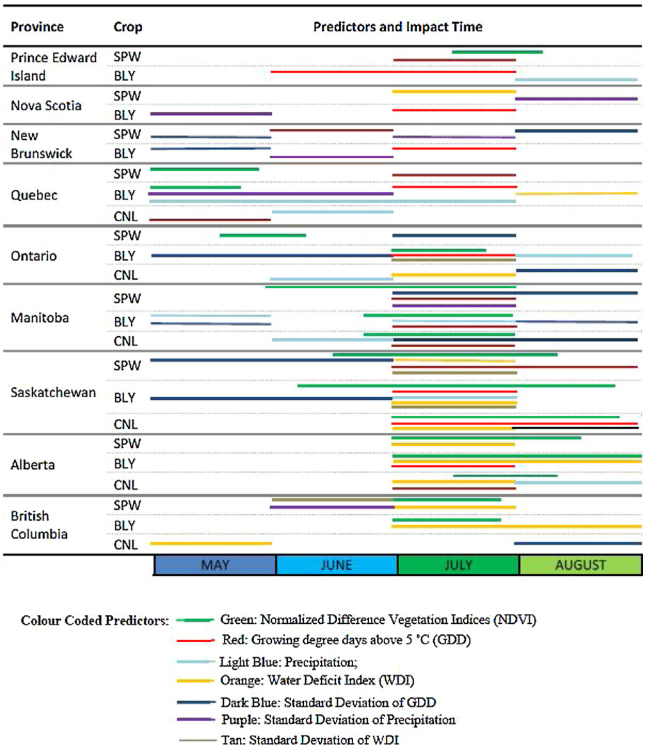
The coefficient of determination (R^2) at CAR level during model calibration showed very distinct regional patterns (Fig. 5). For spring wheat and canola, the R^2 values were higher in BC's Peace River region (CAR 5908), most CARs in AB and SK than in other parts of the country. These regions with higher R^2 had a good coverage of climate stations (Fig. 1) and higher percentages of crop coverage in their agricultural land (Fig. 1 and Table 1). The R^2 values for barley were not as high as those for spring wheat and canola in most of the CARs, revealing the need for constructing crop specific predictors for each crop. The R^2 for all reporting units ranged from 0.30 to 0.90, 0.11 to 0.88 and 0.34 to 0.86, and their median values were 0.66, 0.51 and 0.67 for spring wheat, barley and canola, respectively. The large regional variation in R^2 is likely attributed to spatial variations in climate stations, crop coverage, yield controlling factors identified by current modelling algorithms, etc.

3.3. Selected predictors at CAR level

Although some regional similarities were apparent, the selected predictors generally differed among CARs, hence, the provinces with more reporting units have more CAR level predictors (Table 2).

Crop yields in BC were modelled using two reporting units: Peace River region (CAR 5908) and rest of the BC interior. The Peace River region contributed 88%, 71% and 95% of the provincial total harvested area for spring wheat, barley and canola, respectively. The NDVI from end of June to late-July was selected as a dominant predictor for both spring wheat and barley but not for canola in CAR 5908. The NDVI was seldom selected in other parts of BC as the crops were much scattered and the NDVI values were likely contaminated by noise from other land covers. The selected climate predictors included water related indices (e.g. July WDI and

Table 2
The most frequently selected yield predictors at CAR level for spring wheat (SPW), barley (BLY) and canola (CNL) as listed by provinces.



standard deviation of precipitation in June) for spring wheat and barley and the standard deviation of temperature in August for Canola.

There were 40 CARs from the three Prairie provinces (AB, SK and MB) that contributed approximately 99%, 92% and 99% of the national total harvested area of spring wheat, barley and canola, respectively (Table 1). The predictors in these provinces included almost all variables from NDVI and climate related indices during the growing season (Table 2), largely because of the vast geographical coverage and the diversity in climate, soil and crop managements among the 40 CARs (Mkhabela et al., 2011). In spite of the diversity, there were some regional similarities as well. For example, the water deficit index (WDI) of July was selected as a yield predictor in 18 and 15 CARs out of 28 CARs in AB and SK for spring wheat and barley. Other frequently selected predictors for spring wheat and barley in AB and SK were the growing degree days (GDDs) in July, NDVI from late June to mid-August and the standard deviation of temperature in the early part of the growing season (May and June). Both GDD and WDI in July were selected as predictors for canola yield in 12 out of the 28 CARs in AB and SK. Other top predictors for canola in AB and SK included the simulated crop seeding dates (not shown in Table 2), the standard deviation of temperature in August, precipitation and the water deficit index of July. Only in six out of 28 CARs in AB and SK, was NDVI selected as a predictor for canola yield, which raised the possibility that other remote sensing derived indices (or different temporal aggregation of such indices) might be more relevant. In MB, the dominant predictors were temperature related. Out of the total 12 CARs in MB, seven CARs selected the standard deviation of temperatureas a predictor for spring wheat, seven CARs selected July GDD as a predictor for barley and six CARs selected either standard deviation of temperature in July or Augusts as predictors for canola. The NDVI from late June to late July was also frequently selected as a predictor.

The provinces of QC and ON grow about 2% of spring wheat, 6% of barley and less than 1% of canola of the national total harvested area on average. Both spring wheat and canola were reported at provincial level while barley was reported at CAR level (Table 1). The yield variations were relatively low except for a few CARs in QC for barley (Fig. 4). The correlations of crop yield with climate and NDVI indices were also low probably because of the favourable climate conditions, better crop management practices, such as field drainage systems that exist in these provinces, and sparsely distributed crop area. July heat conditions (e.g. GDD and the Std value of temperature) affected the yield of spring wheat in ON and QC. The most frequently selected predictors for barley in ON and QC included GDD of July. Precipitation of June and Standard deviation of precipitation in May implying that the barley yield was impacted by water conditions in early season and temperature conditions in mid-season. The yield for canola responded to the water stress (WDI) in July and the standard deviation of temperature in August and a few other early predictors, and this conformed with the findings from other studies (Aksouh-Harradj et al., 2006; Kutcher et al., 2010).

The percentage of harvest areas of wheat, barley and canola from Atlantic Canada (PE, NS and NB) was low compared to other provinces (Table 1). The yield model correlations were low in this region (Fig. 5); however, the relative contribution of these provinces to the national production is small. The low correlation in this region may be attributed to the relative large error in the surveyed yields (Statistics Canada, 2014) and of the relative small crop area coverage.

3.4. *Yield forecasting skill according to the lead time*

A distinct feature of the ICCYF is that, it generates in-season yield forecasts well before the end of the growing season and provides a probability distribution of the forecasted yields. Based on the current near real time data flow (Fig. 2), yield forecasts can be generated around the mid-point of the forecast month using all of the observed data from the start of the growing season until the last day of the previous month and the model estimated inputs for the remainder of the growing season. If the selected yield model contains predictors from all time periods of the growing season, the forecasted yield median would draw closer to the actual yield and the reliability range (i.e. RFCI 80) would converge as the forecast time advances towards the end of the growing season and more observations become available. However, with the current model selection algorithm, a yield model may only contain predictors at a few critical stages; thus, the forecast error and reliability converging trend could deviate from the ideal expectation. For example, if only predictors before June were selected in a CAR, the new observations from July and August would have no effect on the forecasted yield, thus, the forecasted yield and the reliability range will remain the same in forecasts obtained after mid-July.

The overall trends of change in forecast error (MAPE) and reliability range (RFCI 80) over the four tested forecasts (June–September) were consistent across the three crop types (Fig. 6). The box plots shown were obtained from all the forecasts for all the CARs over the entire period of study during the LOOCV test. At each forecast point, the observed data after the last day of previous month were replaced by the model generated data using random forecasts algorithm (Fig. 2) to mimic the near real time forecast situation. Both the MAPEs (Fig. 6, left panels) and the RFCI 80s (Fig. 6, right panels) decreased significantly from July forecasts to August forecasts for all three crops, but the improvement between any other two consecutive forecasts was small. This indicates that the predictors in July are critical to all the three crops and a skilful forecast is likely achievable at mid-August.

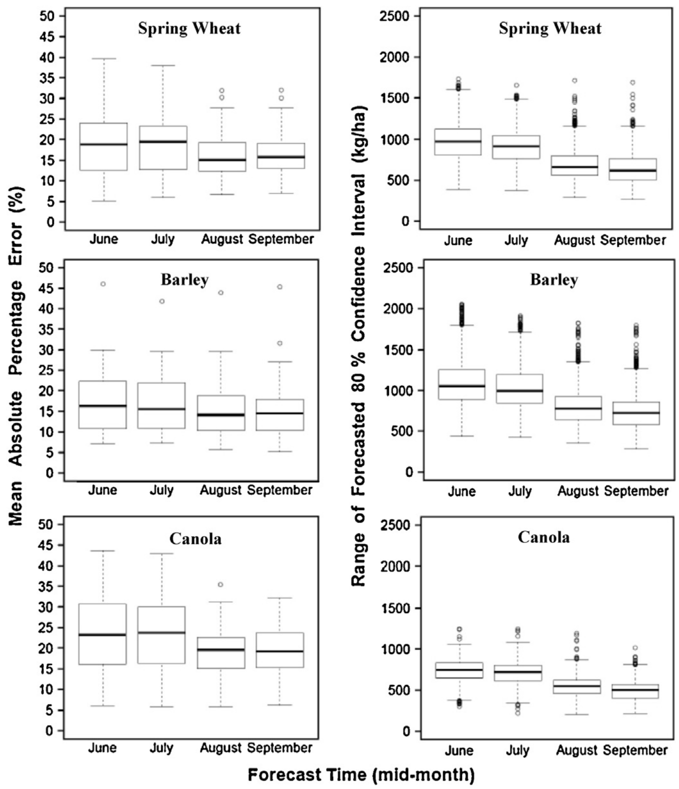


Fig. 6. Boxplots of the mean absolute percentage error (MAPE, left) and the range of forecasted 80% confidence interval (right) for mid-June–September forecasts at the census agricultural region (CAR) scale. The plots included results from all years for all CARs during the leave-one-out-cross-validation. The thick horizontal lines represent the median, the top and bottom of the box represent the 25th and 75th percentiles and the whiskers are 1.5 times of the box height towards upper and lower from the median. Open circles are outliers.

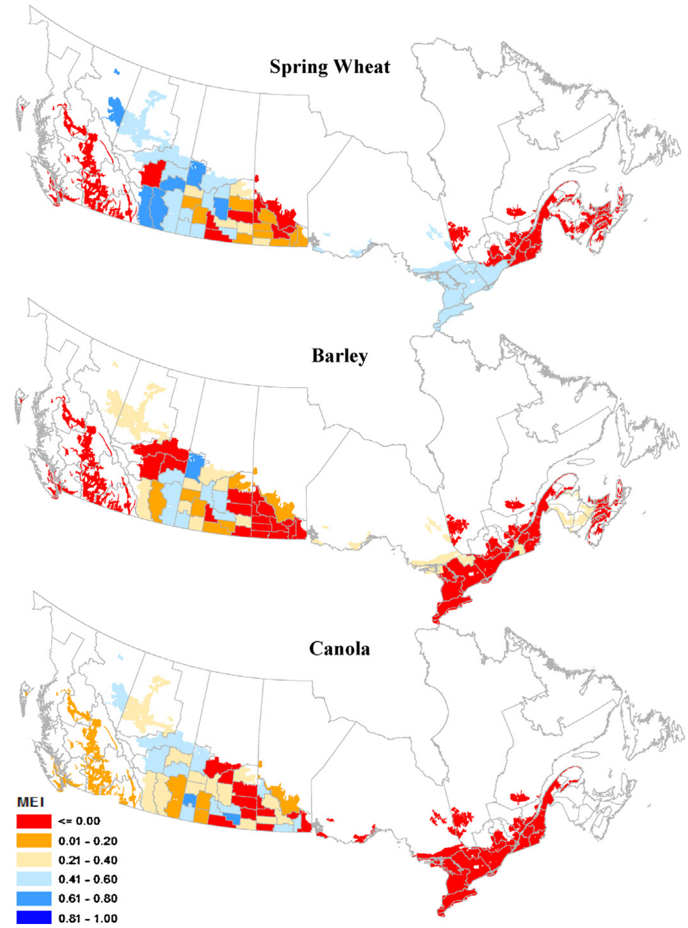


Fig. 7. Distribution of model efficiency index (MEI) of the yield models at the census agricultural region (CAR) scale across Canada for the three tested crops. MEI values were evaluated with September forecasts obtained by the LOOCV testing. Only the crop land within a given CAR is used to map the results. Areas in red (MEI < 0) indicates where mean yield is a better estimation than the model outputs. MEI towards 1 (blue) indicates a perfect model.

3.5. Spatial distribution of CAR level forecasting skill (MEI) and quality (MAPE) across Canada

To reduce redundancy, only results from September forecasts were shown for the spatial analysis of forecast skill (MEI) and quality (MAPE) statistics. The September forecasts integrated all available information of the forecast years, thus, represent the best forecast results that were achieved by ICCYF during the LOOCV tests, although their differences from August forecasts were insignificant (Fig. 6). The spatial distribution of the MEI identified the regions where the current ICCYF models need significant improvements (Fig. 7). The percentage of CARs with positive MEI was 70%, 43% and 70% for spring wheat, barley and canola, respectively. The majority of CARs with negative MEI values were located outside the Prairie region, where the harvest area was small. Among the three Prairie provinces, models in AB performed best, while models in MB performed the worst. Possible explanations are: (1) AB has the highest density of climate stations within the crop lands (Fig. 2), thus, the regional representation of the climate at the CAR scale is better than other regions where climate stations are scarce; (2) all three crops are grown extensively throughout AB and SK, not in some MB CARs; the current non-crop specific crop mask used to filter the NDVI pixels is thus, more representative in AB and SK CARs than in MB CARs; (3) many studies (e.g. Campbell et al., 2002; Kutcher et al., 2010) showed that the crop yield variations in AB and SK are largely determined by water availability during the season, which were well captured by the ICCYF selected predictors (e.g. WDI of July), while the factors causing yield variation in MB are much more complicated (Entz et al., 2001; Qian et al., 2009a).

Not surprisingly, among the three tested crops, the models performed best for spring wheat, followed by canola and barley. Indeed, the current parameters used for phenological development calculation to scale the crop water use rate have only been tested with spring wheat (Robertson, 1968). Crop specific parameterisations for barley and canola are still under development.

The spatial distribution of forecast quality index MAPE (Fig. 8) was different from the distribution of forecast skill index MEI (Fig. 7). Forecasted relative error was much larger in Western Canada (BC, AB, SK and MB) than in Eastern Canada (ON, QC, NB, PE and NS). This is because the historical yield variation is much larger in Western Canada than in Eastern Canada (Fig. 4). While the MAPE will guide the forecast users assessing their risk level as part of their decision making process, the MEI is more useful to the forecaster in terms of pointing out where model improvements are needed. Although the forecasting errors were higher in many Prairie CARs than in Eastern Canada CARs, the forecasts were more valuable in the Prairie CARs as the forecasted errors were mostly smaller than the historical yield variations. Whereas, smaller forecasting errors were achieved at CARs in Eastern provinces (Fig. 8), the forecast results with the ICCYF were still inferior to using the historical mean as forecasts (Fig. 7). The average MAPE at CAR level for all CARs across Canada were 16%, 15% and 19% for spring wheat, barley and canola, respectively, which were below the mean historical CV of 21%, 17% and 25%, correspondingly.

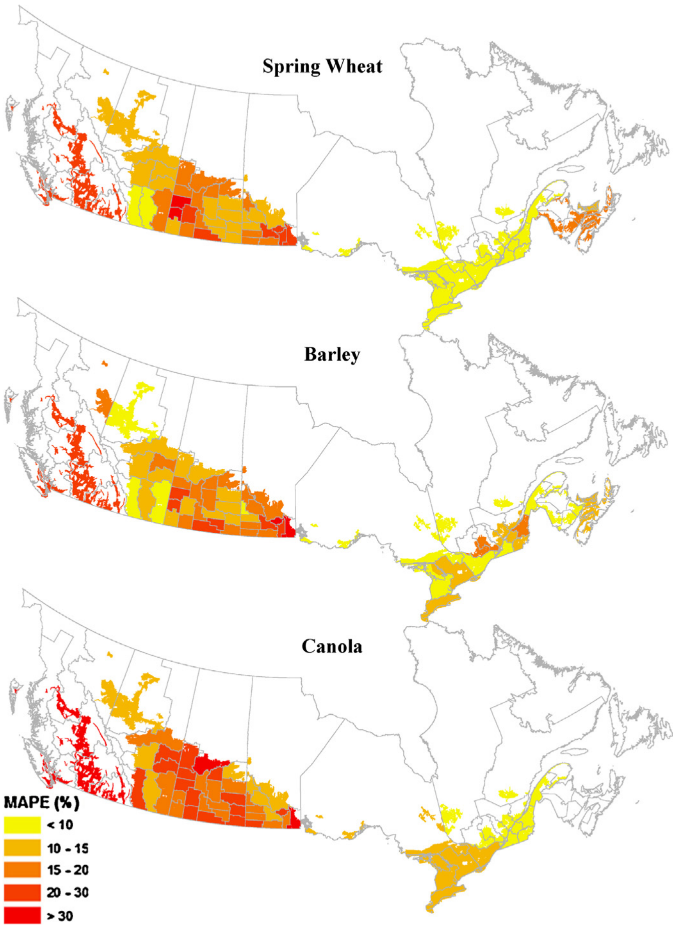


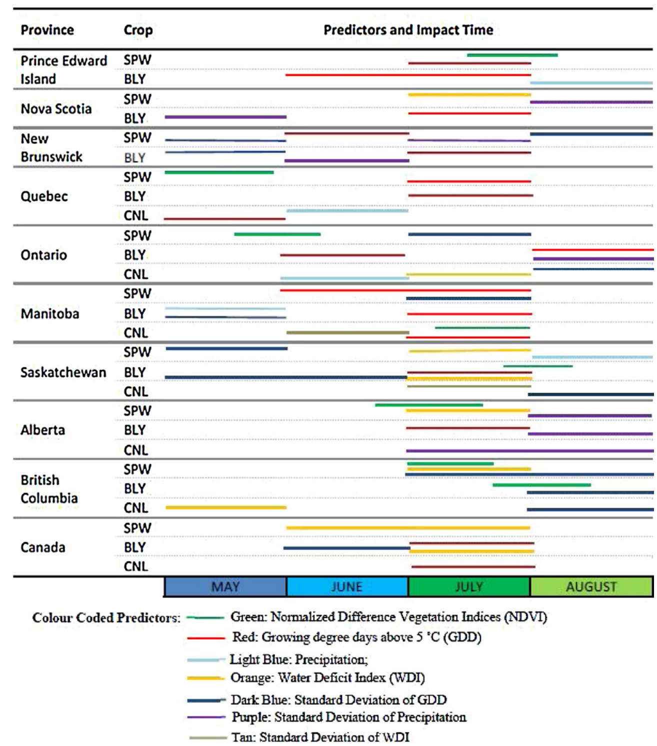
Fig. 8. Distribution of mean absolute percentage error (MAPE) of the yield models at the census agricultural region (CAR) scale across Canada for the three tested crops. MAPE values were evaluated with September forecasts obtained by the leave-one-out cross-validation. Only the crop land within a given CAR is used to map the results.

3.6. Aggregating yield forecastsat provincial and national scales

The results from both methods for the yield forecasting at provincial and national scales, i.e. the YAM and the PAM as described in Section 2.4, were compared. The predictors for YAM- and PAM-based models were those listed in Tables 2 and 3, respectively. The predictors selected at provincial and national scales captured some of the key predictors at CAR scales, such as the WDI of July for spring wheat and barley and GDD of July and standard deviation of daily temperature in August for canola. However, the majority of the CAR scale predictors were lost during the predictor aggregation, especially the NDVI indices from the Prairie provinces.

For all the three performance measures (i.e. R^2, MEI and MAPE), the YAM method performed better than the PAM method for spring wheat and barley at the national scale and at all the Western provinces except MB, where the PAM performed as good as YAM or slightly better (Tables 4–6). Regarding the forecasts of canola yield, the YAM method was superior to the PAM in MB and BC. For all the three crops, from central and Atlantic Canada, the two methods did not differ significantly. It is also noted that the YAM was superior to the PAM in regions where credible models at the CAR scale were achieved (Tables 4–6). Based on this result, the YAM is the preferred aggregation method over PAM especially when smaller scale data are available and reliable. Overall, the YAM forecasted yield (prediction) followed the surveyed yield trend and variation reasonably well at the national level and at most provinces (Fig. 9), especially in those provinces with a high percentage of harvested area (e.g. AB, SK and MB). However, the current model seemed weak in forecasting extreme yields, e.g. the extremely low spring wheat and barley yields in 1988 at BC, SK, MB and Canada. Lacking sensitivity to extreme events is a common limitationof most statistical models.

Table 3
The most frequently selected yield predictors at provincial and national levels for spring wheat (SPW), barley (BLY) and canola (CNL).



4. Discussion

The Integrated Canadian Crop Yield Forecaster (ICCYF) was evaluated using multiple model performance measures with a

Table 4
Bravais and Pearson coefficient of determination (R^2) of the forecasted provincial and national yields of the three crops obtained during the leave-one-out cross-validation.

Province	Spring wheat		Barley		Canola	
	YAM[a]	PAM[b]	YAM	PAM	YAM	PAM
PE	0.02	0.02	0.00	0.00	n.a	n.a
NS	0.03	0.03	0.10	0.10	n.a	n.a
NB	0.00	0.00	0.35^{*}	0.35^{*}	n.a	n.a
QC	0.15	0.15	0.00	0.18	0.08	0.08
ON	0.46^{**}	0.46^{**}	0.02	0.03	0.12	0.12
MB	0.19	0.20	0.02	0.04	0.47^{**}	0.23
SK	0.66^{**}	0.52^{*}	0.44^{**}	0.31^{*}	0.39^{*}	0.44^{**}
AB	0.76^{**}	0.61^{**}	0.25	0.23	0.59^{**}	0.61^{**}
BC	0.67^{**}	0.52^{**}	0.33^{*}	0.01	0.52^{**}	0.31
Canada	0.75^{**}	0.63^{**}	0.53^{**}	0.32^{*}	0.55^{**}	0.63^{**}

n.a: not applicable.
[a] Yield-aggregating method (YAM): this method models the yields at census agricultural region (CAR) level and aggregates up the results to provincial and national scales using the surveyed crop area for each CAR.
[b] Predictor-aggregating method (PAM): all the predictors are firstly averaged to provincial scale using the surveyed crop area. Then yields are directly forecasted at the provincial scale.
* *t*-test passed $P < 0.05$ significant level.
** *t*-test passed $P < 0.01$ significant level.

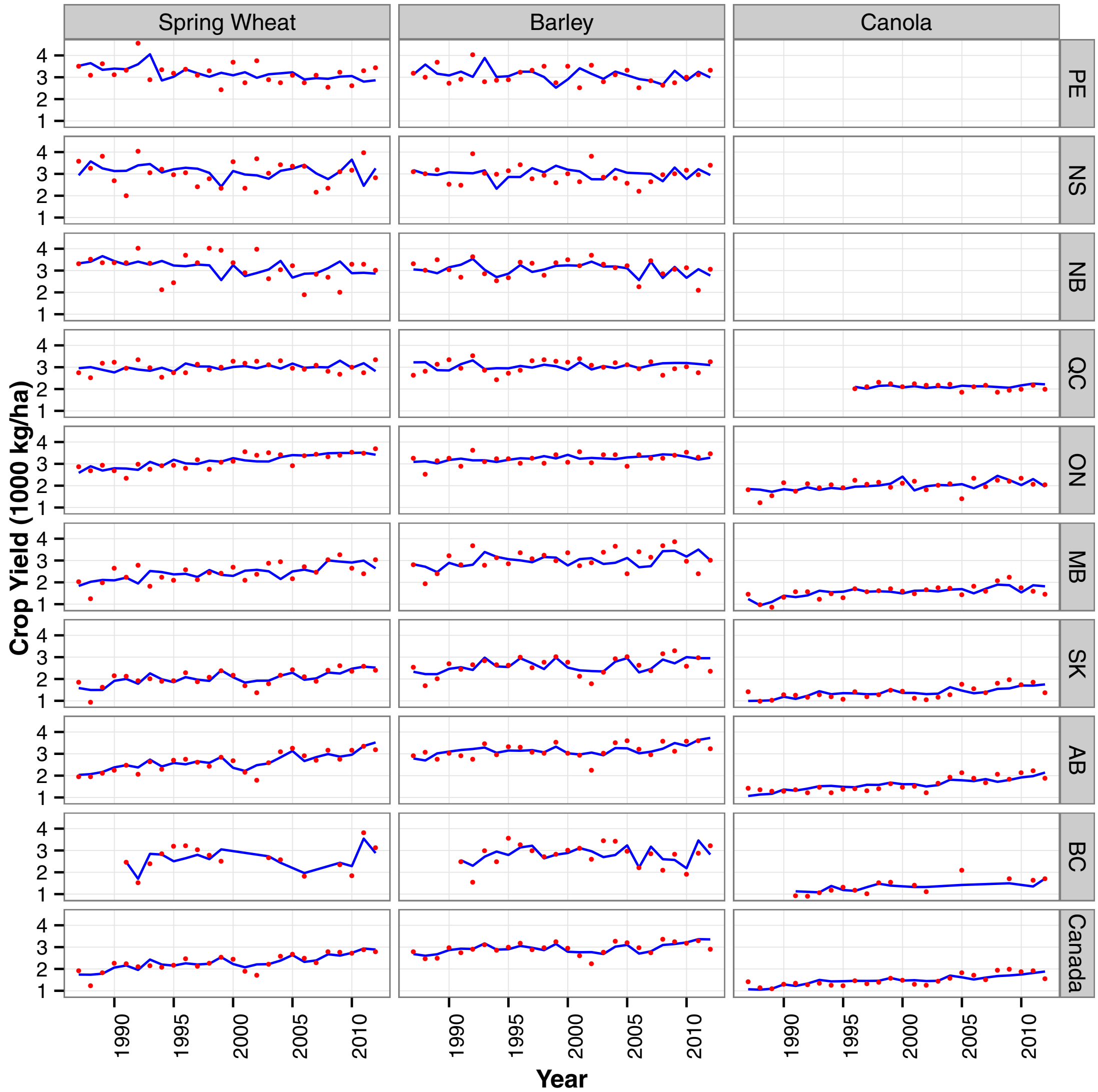


Fig. 9. Forecasted provincial and national yields (lines) vs. surveyed yields (closed circles) of spring wheat, barley and canola using ICCYF during a LOOCV test. BC, AB, SK, MB, ON, QC, NB, NS and PE refer to the province of British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Quebec, New Brunswick, Nova Scotia and Prince Edward Island, respectively. Data are not available for canola yield in PE, NB and NS (blank panels).

Table 5
Model efficiency index (MEI) of the forecasted provincial and national yields for the three crops obtained during the leave-one-out cross-validation.

Province	Spring Wheat		Barley		Canola	
	YAM[a]	PAM[b]	YAM	PAM	YAM	PAM
PE	−0.24	−0.24	−0.49	−0.49	n.a	n.a
NS	−0.12	−0.12	−0.70	−0.70	n.a	n.a
NB	−0.16	−0.16	0.35	0.35	n.a	n.a
QC	−0.67	−0.67	−0.25	0.12	−0.19	−0.19
ON	0.45	0.45	−0.08	−0.59	−0.01	−0.01
MB	0.12	0.10	−0.18	−0.10	0.46	0.09
SK	0.66	0.47	0.44	0.25	0.36	0.39
AB	0.76	0.61	0.21	0.12	0.59	0.60
BC	0.66	0.50	0.33	−0.13	0.46	0.27
Canada	0.74	0.62	0.53	0.20	0.55	0.63

n.a: not applicable.
[a] Yield-aggregating method (YAM): this method models the yields at CAR level and aggregates up the results to provincial and national scales using the surveyed crop area for each CAR.
[b] Predictor-aggregating method (PAM): all the predictors are firstly averaged to provincial scale using the surveyed crop area. Then yields are directly forecasted at the provincial scale.

leave-one-out-cross-validation procedure during 1987–2012 for three major field crops (spring wheat, barley and canola) at three regional scales across the Canadian agricultural landscape. At the CAR level, the model performed better in Alberta and Saskatchewan than other parts of the country, mostly because of the good coverage of climate stations and a high percentage area in cultivated cropland. At the national scale and at most provinces where the percentage of the crop area over the total agricultural land was high the ICCYF also achieved satisfactory performances irrespective of the crops involved. In Western Canada (encompassing the three major crop production provinces) the MEIs obtained at the provincial scale were all above zero, while most ICCYF yield models in central Canada and Atlantic provinces require significant improvements (though the contributions of these provinces to the national yield forecasts were relatively small due to their percentage of harvested area).

The current ICCYF covers a broad spectrum of variables (e.g. GDD, P, WDI, SWA, CSD, NDVI and some of the standard derivations associated with these variables) aspotential yield predictors. Although some of the selected predictors are well-known biophysical controls of crop yield (e.g. WDI of July in the arid and semi-arid CARs in the Prairies), the selection of others at certain time intervals

Table 6
Mean absolute percentage error (MAPE) of the forecasted provincial and national yields for the three crops obtained during the leave-one-out cross-validation.

Province	Spring Wheat			Barley			Canola		
	CV[a] (%)	YAM[b] (%)	PAM[c] (%)	CV (%)	YAM (%)	PAM (%)	CV (%)	YAM (%)	PAM (%)
PE	14.0	12.4	12.4	12.3	11.2	11.2	n.a	n.a	n.a
NS	18.3	16.1	16.1	13.1	14.4	14.4	n.a	n.a	n.a
NB	18.7	16.5	16.5	13.0	8.5	8.5	n.a	n.a	n.a
QC	8.2	8.5	8.5	9.0	8.8	6.6	6.5	6.1	6.1
ON	11.1	7.0	7.0	7.5	6.6	8.1	13.6	11.3	11.3
MB	18.1	14.4	14.4	15.0	13.2	13.7	18.4	11.6	15.4
SK	18.5	10.0	11.0	15.3	9.7	11.0	19.9	12.3	12.0
AB	17.0	6.6	8.0	10.3	7.4	7.7	20.0	11.1	9.9
BC	22.3	10.3	14.4	18.3	12.5	16.8	25.4	13.6	15.2
Canada	16.9	7.3	8.6	9.6	5.3	6.9	17.3	9.1	8.5

n.a: not applicable.
[a] CV: coefficient of variation of provincial scale surveyed yield during 1987–2012.
[b] Yield-aggregating method (YAM): this method models the yields at CAR level and aggregates up the results to provincial and national scales using the surveyed crop area for each CAR.
[c] Predictor-aggregating method (PAM): all the predictors are firstly averaged to provincial scale using the surveyed crop area. Then yields are directly forecasted at the provincial scale.

(e.g. NDVI in May and early June, standard deviation of temperature in August) lacks obvious biophysical explanation. For example, the selection of certain variables as predictors at non critical times during the crop calendar was most likely a statistical coincidence. Future efforts will be spent on exploring biophysical based variables as potential predictors, especially those indices related to extreme climates. This will ultimately improve the yield forecasting in extreme years.

In the current set up of the ICCYF, a fixed time interval to aggregate the agroclimatic (monthly) and NDVI (3-weekly) indices is used. As a result, the derived variables are not based on phenological development. Studies have shown that crop yield response to climate, soil water and other environmental factors is closely related to crop phenological stages (Aksouh-Harradj et al., 2006; Qian et al., 2009a; Kutcher et al., 2010). The next iteration of the ICCYF will involve a phenology-based predictor-aggregation algorithm. Efforts to quantify the crop phenology using remote sensing inputs will also be explored (e.g. Bolton and Friedl, 2013). Furthermore, the wheat crop was used as the reference crop in the VSMB model with the assumption that the response of barley and canola crops to the soil and atmospheric environments is similar to that of spring wheat. As such, the crop and soil water extraction coefficients for spring wheat were applied to barley and canola. The results of this study showed that although this assumption was a reasonable one, results for spring wheat were the best in comparison to the other two crops. As such, a future development activity of the ICCYF is to build crop-specific and phenology-based indices as predictors.

The fact that the model performed unfavourably in regions where the crop coverage was low was possibly exacerbated by using a common vegetation mask for all crops instead of a crop specific land cover mask. Research efforts are now directed at developing crop specific land cover masks across the Canadian agricultural landscape that will be applied in next generation of the ICCYF. This will allow the accurate delineation of the crop specific area and more adequately define the regional representative predictors. Moreover, investigating satellite remote sensing sensors with different spatial and temporal resolutions will be an asset for improving the ICCYF performances. The 1 km spatial resolution of AVHRR products used in this study were preferred because of its longer time series and the sample size requirement in calibrating the yield model using historical data. Moderate spatial resolution products such that derived from the MODerate-resolution Imaging Spectroradiometer (MODIS, 250 m resolution) will be explored in the future to retrieve the vegetation index. Indeed, such data could be useful when the sub-CAR scale spatial variation of yield needs to be analysed (e.g. Johnson, 2014). Some other remote sensing indices, such as enhanced vegetation index (EVI) are also under consideration for potential yield predictors, especially in those regions where AVHRR NDVI did not perform well (Chen et al., 2006).

The smallest working scale in this study was the CAR, which is selected due to the fact that they are the units at which historical yield data is publicly available. The results showed that predicting the yields at a smaller scale and aggregating them to a larger scale (i.e. YAM) produced better results than directly predicting yield at the larger scale using aggregated predictors (PAM). Apparently, predictors that contribute significantly to the total variance in yield will have physical meaning at the local level than at the regional level. Recognizing that the CAR is a coarse unit and it normally contains many different soil and climate zones (Mkhabela et al., 2011) and thus, may have different yield response relationships, developing predictive models at finer scales improves both the forecast quality and expand forecasts to more localized user groups. However, this is only feasible if yield data at sub-CAR scales is available to calibrate the model. Efforts are currently underway to obtain historical yield data at finer scales through alternative sources (e.g. crop insurance company, provincial agricultural services, etc.).

The ICCYF is currently calibrated to historical yield variation to generate forecasts. Due to the statistical nature of ICCYF it tends to underestimate the extremely high crop yields and overestimated the extremely low yields (Fig. 9). Process-based models are a good tool to capture those extreme years because responses to environmental predictors mainly rely on the crop physiology rather than historical information. With the efforts to reduce the input and parameterization requirements that hinders the regional application of process-based models (e.g. Folberth et al., 2012; Basso et al., 2013), using them into regional yield forecasting tools, such as the ICCYF becomes increasingly possible. We have embarked on an effort to embed one or more biophysical models [e.g. APSIM (Keating et al., 2003) and WOFOST (Diepen et al., 1989)] into the ICCYF, especially at sub-CAR scales. This will also provide further exploration of predictors that better represent the accumulated effects over the growing season including the extreme impacts.

5. Conclusions

The assessment of the Integrated Canadian Crop Yield Forecaster (ICCYF) for three major crops at different spatial and temporal scales showed that the forecast reliability is improved over the cropping season when more near real time data became available. A skilful forecast using the ICCYF could be expected around mid-August. This gives a lead time of about 1 month before harvest and about 3–4 months of lead time before the official final release of Statistics Canada's survey results, which is often made publicly available in December. As the official yields in this study are from the final release of the four annual yield surveys conducted by Statistics Canada, it was hypothesized that the forecasts from the ICCYF are not significantly different from the Statistics Canada's survey reports. This hypothesis has implications on the continued use of the survey method for reporting crop yields across Canada. In spite of the regional differences in the performance of the ICCYF across Canada, the results were within the range or better than the observed variability of yields as reported by Statistics Canada. It is conceivable therefore, that given the dwindling resources provided for conducting four surveys of crop yields during the growing season by Statistics Canada, the ICCYF could potentially complement some of the early season surveys, especially when the gaps in data and model as discussed in this study are addressed.

Acknowledgements

This study was supported by three Agriculture and Agri-Food Canada (AAFC) funded projects (1) "Climate Adaptation for Resilience in Agriculture (CARA)", (2) "the Sustainable Agriculture Environmental Systems (SAGES) – Integrated Earth Observation Information on Crops and Soils for Agri-Environmental Monitoring in Canada" and (3) "Integrated Canadian Crop Yield Forecasting (ICCYF) across the agriculture Landscape of Canada". Parts of the data used in this study were provided by partners in Statistics Canada and Environment Canada. The authors would like to thank the following colleagues and students for their support and contributions during various stages of this study: Robert Armstrong, Catherine Champagne, Patrick Cherneski, James Ehrman, Xiaoyuan Geng, Evgeniy Gorelov, Allan Howard, Ian Jarvis, David Kroetsch, Zhong-Yao Lin, Heather McNairn, Dongzhi Qi, Richard Rieger, Patrick Rollin, Jiali Shang, Souleymane Toure, Lawrence Townley-Smith, Jonathan Van Dusen, David Waldner and David Zamar. We also thank the two anonymous reviewers for providing thoughtful comments and suggestions.

References

Agriculture and Agri-Food Canada, 2005. Land Cover for Agricultural Regions of Canada. circa 2000. URL: <http://data.gc.ca/data/en/dataset/f5ded3b0-a5b4-4599-95d6-d853a825792b> (accessed on March 2015).

Akinremi, O.O., McGinn, S.M., Barr, A.G., 1996. Simulation of soil moisture and other components of the hydrological cycle using a water budget approach. Can. J. Soil Sci. 76 (2), 133–142.

Aksouh-Harradj, N.M., Campbell, L.C., Mailer, R.J., 2006. Canola response to high and moderately high temperature stresses during seed maturation. Can. J. Plant Sci. 86 (4), 967–980.

Atzberger, C., 2013. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens. 5 (2), 949–981.

Baier, W., Boisvert, J.B., Dyer, J.A., 2000. The versatile soil moisture budget (VSMB) reference manual [computer software], ECORC contribution no. 1553. In: Agriculture and Agri-Food Canada. Eastern Cereal and Oilseed Research Centre, Ottawa, ON, Canada, pp. A1–D4.

Basso, B., Cammarano, D., Carfagna, E., 2013. Review of crop yield forecasting methods and early warning systems. In: Report Presented to First Meeting of the Scientific Advisory Committee of the Gloal Strategy to Improve Agricultural and Rural Statistics. FAO, Headquarters, Rome, Italy, 18–19 July.

Basso, B., Chou, T.Y., Chen, C., Yeh, M., 2012. iSalus: new web based spatial systems for simulating crop yield and environmental impact. In: Proceedings International Conference on Precision Agriculturee, Indianapolis.

Bolton, D., Friedl, M., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric. For. Meteorol. 173, 74–84.

Boken, V.K., Shaykewich, C.F., 2002. Improving an operational wheat yield model using phenological phase-based normalized difference vegetation index. Int. J. Remote Sens. 23 (20), 4155–4168.

Bornn, L., Zidek, J.V., 2012. Efficient stabilization of crop yield prediction in the Canadian Prairies. For. Meteorol. 152 (0), 223–232.

Bullock, P.R., 1992. Operational estimates of western Canadian grain production using NOAA AVHRR LAC data. Can. J. Remote Sens. 18 (1), 23–28.

Cammarano, D., 2009. Spatial Integration of Remote Sensing and Crop Simulation Modelling for Wheat Nitrogen Management. University of Melbourne, Victoria, Australia, 422 pp. (Ph.D. thesis).

Campbell, C.A., Selles, F., Zentner, R.P., McConkey, B.G., Brandt, S.A., McKenzie, R.C., 1997. Regression model for predicting yield of hard red spring wheat grown on stubble in the semiarid prairie. Can. J. Plant Sci. 77 (1), 43–52.

Campbell, C.A., Selles, F., Gameda, S., Blomert, B., Wall, D., 2002. Production of annual crops on the Canadian Prairies: trends during 1976–1998. Can. J. Soil Sci. 82 (1), 45–57.

Canadian System of Soil Classification Working Group, 1998. The Canadian System of Soil Classification, Third ed. Agriculture and Agri-Food Canada, 1646, 187 pp. (revised).

Chen, P.Y., Fedosejevs, G., Tiscareño-López, M., Arnold, J.G., 2006. Assessment of MODIS–EVI, MODIS–NDVI and VEGETATION–NDVI composite data using agricultural measurements: an example at corn fields in western Mexico. Environ. Monit. Assess. 3, 69–82.

Chipanshi, A.C., Ripley, E.A., Lawford, R.G., 1999. Large-scale simulation of wheat yields in a semi-arid environment using a crop-growth model. Agric. Syst. 59 (1), 57–66.

Chipanshi, A.C., Zhang, Y., Newlands, N.K., Hill, H., Zamar, D.S., 2012. Canadian crop yield forecaster (CCYF)–a GIS and statistical integration of agroclimates and remote sensing information. In: Proceedings of the Workshop on the Application of Remote Sensing and GIS Technology on Crops Productivity among APEC Economies, Beijing, PR China, p. 14.

De Wit, A., 2007. Regional Crop Yield Forecasting Using Probabilistic Crop Growth Modelling and Remote Sensing Data Assimilation. Wageningen University, Wageningen, 182 pp. (Ph.D. thesis).

Diepen, C.A., Wolf, J., Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. Soil Use Manage. 5 (1), 16–24.

Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. Remote Sens. Environ. 92 (4), 548–559.

Dowd, M., 2006. A sequential Monte Carlo approach for marine ecological prediction. Environmetrics 17 (5), 435–455.

Efron, B., 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. J. Am. Statist. Assoc. 78 (382), 316–331.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32 (2), 407–499.

Environment Canada, 2014. climate data online access portal. URL: <http://climate.weather.gc.ca/> (accessed on March 2015).

FAO, 2014. Global Information and Early Warning System (GIEWS) on food and agriculture, Trade and Markets Division, FAO. URL: <http://www.fao.org/Giews/english/index.htm> (accessed on March 2015).

Folberth, C., Gaiser, T., Abbaspour, K.C., Schulin, R., Yang, H., 2012. Regionalization of a large-scale crop growth model for sub-Saharan Africa: model setup, evaluation, and estimation of maize yields. Agric. Ecosyst. Environ. 151, 21–33.

Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: a web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Comput. Electron. Agric. 84 (0), 111–123.

Hochheim, K.P., Barber, D.G., 1998. Spring wheat yield estimation for Western Canada using NOAA NDVI data. Can. J. Remote Sens. 24, 17–27.

Huang, J., Han, D., 2014. Meta-analysis of influential factors on crop yield estimation by remote sensing. Int. J. Remote Sens. 35 (6), 2267–2295.

Entz, M.H., Guilford, R., Gulden, R., 2001. Crop yield and soil nutrient status on 14 organic farms in the eastern portion of the northern great plains. Can. J. Plant Sci. 81, 351–354.

Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141 (0), 116–128.

Joint Research Centre, 2012. Data distribution-MARS crop yield forecasting system (MCYFS). European Commission. URL: <http://mars.jrc.ec.europa.eu/mars/About-us/AGRI4CAST/Data-distribution> (accessed on March 2015).

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18 (3–4), 267–288.

Khan, J., Aelst, S.V., Zamar, R., 2010. Fast robust estimation of prediction error based on resampling. Comput. Stat. Data Anal. 54, 3121–3130.

Khan, J.A., Van Aelst, S., Zamar, R.H., 2007. Robust linear model selection based on least angle regression. J. Am. Stat. Assoc. 102 (480), 1289–1299.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.

Kutcher, H.R., Warland, J.S., Brandt, S.A., 2010. Temperature and precipitation effects on canola yields in Saskatchewan, Canada. Agric. For. Meteorol. 150 (2), 161–165.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R. News 2 (3), 18–22.

Mkhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y., 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric. For. Meteorol. 151 (3), 385–393.

Moulin, A.P., Beckie, H.J., 1993. Evaluation of the CERES and EPIC models for predicting spring wheat grain yield over time. Can. J. Plant Sci. 73 (3), 713–719.

Murphy, A.H., 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather Forecast. 8 (2), 281–293.

NASA, 2013. MOD 13 – Gridded Vegetation Indices (NDVI & EVI) [Data]. URL: <http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=13> (acessed on April 2013).

Newlands, N.K., Zamar, D.S., 2012. In-season probabilistic crop yield forecasting – integrating agro-climate, remote-sensing and crop phenology data. In: Proceedings of the 2012 Joint Statistical Meetings (2012 JSM): Statistics – Growing to Serve a Data Dependent Society, San Diego, CA, USA, pp. 1–10.

Newlands, N.K., Zamar, D., Kouadio, L., Zhang, Y., Chipanshi, A., Potgieter, A., Toure, S., Hill, H.S.J., 2014. An integrated model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. Front. Environ. Sci. 2, 17, http://dx.doi.org/10.3389/fenvs.2014.00017.

Nikolova, S., Bruce, S., Randall, L., Barrett, G., Ritman, K., Nicholson, M., 2012. Using remote sensing data and crop modelling to improve crop production forecasting: a scoping study. In: Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES). ABARES Technical Report 12.3, Canberra, Australia, p. 23.

NOAA, 2013. GVI – Normalized Difference Vegetation Index. Office of Satellite and product operations, The National Oceanic and Atmospheric Administration, USA. URL: <http://www.ospo.noaa.gov/Products/land/gvi/NDVI.html> (acessed on March 2015).

Phillips, D., 1990. The Climate of Canada. Catalogue No. En56-1/1990E. Ottawa Minister of Supply and Services of Canada. URL: <http://climate.weather.gc.ca/prods_servs/historical_publications_e.html> (acessed on March 2015).

Potgieter, A.B., Everingham, Y.L., Hammer, G.L., 2003. On measuring quality of a probabilistic commodity forecast for a system that incorporates seasonal climate forecasts. Int. J. Climatol. 23 (10), 1195–1210.

Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8 (1), 26–33.

Qian, B., De Jong, R., Gameda, S., 2009a. Multivariate analysis of water-related agroclimatic factors limiting spring wheat yields on the Canadian Prairies. Eur. J. Agron. 30 (2), 140–150.

Qian, B., De Jong, R., Warren, R., Chipanshi, A., Hill, H., 2009b. Statistical spring wheat yield forecasting for the Canadian prairie provinces. Agric. For. Meteorol. 7, 1022–1031.

Qian, B., De Jong, R., Gameda, S., Huffman, T., Neilsen, D., Desjardins, R., Wang, H., McConkey, B., 2013. Impact of climate change scenarios on Canadian agroclimatic indices. Can. J. Soil Sci. 93, 243–259.

Reichert, G.C., Caissy, D., 2002. A Reliable Crop Condition Assessment Program (CCAP) Incorporating NOAA AVHRR Data, a Geographical Information System, and the Internet, Environmental Systems Research Institute (ESRI) User Conference San Diego, California. URL: <http://www26statcan.ca/ccap-peec/esri2002conf-eng.jsp> (accessed on March 2015).

Ritchie, J.T., Alagarswamy, G., 2002. Overview of crop models for assessment of crop production. In: Doering, O.C.I., Randolph, J.C., Southworth, J., Pfeifer, R.A. (Eds.), Effects of Climate Change and Variability on Agricultural Production Systems. Kluwer Academic Publishing, Dordrecht, The Netherlands, pp. 43–68.

Robertson, G.W., 1968. A biometeorological time scale for a cereal crop involving day and night temperatures and photoperiod. Int. J. Biometeorol. 12 (3), 191–223.

Schut, A.G.T., Stephens, D.J., Stovold, R.G.H., Adams, M., Craig, R.L., 2009. Improved wheat yield and production forecasting with a moisture stress index, AVHRR and MODIS data. Crop Pasture Sci. 60 (1), 60–70.

Shields, J.A., Tarnocai, C., Valentine, K.W.G., Macdonald, K.B., 1991. Soil Landscapes of Canada: Procedures Manual and User Handbook. LRRC Contribution Number 88-29. Agriculture Canada, Land Resource Research Centre, Research Branch, Ottawa, ON, Canada (accessed on March 2015) http://sis.agr.gc.ca/cansis

Statistics Canada, 2010. Definitions, data sources and methods: Satellite image data processing at Statistics Canada for the Crop Condition Assessment Program (CCAP). Statistics Canada, Agriculture Statistics Division, Online documents. URL: <http://www23statcan.gc.ca/imdb-bmdi/document/5177_D1_T9_V1-eng.htm> (accessed on March 2015).

Statistics Canada, 2012. Census Agricultural Regions Boundary Files of the 2011Census of Agriculture. Statistics Canada, Agriculture Statistics Division, URL: <http://www.statcan.gc.ca/pub/92-637-x/92-637-x2011001-eng.htm> (accessed on March 2015).

Statistics Canada, 2013. CANSIM Table 001-0071: estimated areas, yield and production of principal field crops by Small Area Data Regions, in metric and imperial units, annual, 1976 to 2014. Statistics Canada, Agriculture Statistics Division. URL: <http://www5statcan.gc.ca/cansim/a33?RT&9552;TABLE&themeID=2024&spMode=tables&lang=eng> (accessed on March 2015).

Statistics Canada, 2013. CANSIM Table 001-0010 – Estimated areas, yield, production and average farm price of principal field crops, in metric units, annual, 1908 to 2013. Statistics Canada, Agriculture Statistics Division. URL: <http://www5statcan.gc.ca/cansim/a33?RT&9552;TABLE&themeID=2024&spMode=tables&lang=eng> (accessed on March 2015).

Statistics Canada, 2014. Definitions, data sources and methods of Field Crop Reporting Series, Record number: 3401. Statistics Canada, Agriculture Statistics Division. URL: <http://www23statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3401> (accessed on March 2015).

Supit, I., 1997. Predicting national wheat yields using a crop simulation and trend models. Agric. For. Meteorol. 88 (1–4), 199–214.

Szulczewski, W., Żyromski, A., Biniak-Pieróg, M., 2012. New approach in modeling spring wheat yielding based on dry periods. Agric. Water Manage. 103, 105–113.

Thompson, L.M., 1969. Weather and technology in production of corn in US corn belt. Agron. J. 61, 453–456.

USAID, 2014. Famine Early Warning System Network (FEWS NET), United States Agency for International Development (USAID). URL: <http://www.fews.net/> (acessed on March 2015).

USDA, 2012. The yield forecasting program of NASS, SMB staff report number SMB 12-01 United States Department of Agriculture (USDA). URL: <http://www.nass.usda.gov/Publications/Methodology_and_Data_Quality/Advanced_Topics/Yield%20Forecasting%20Program%20of%20NASS.pdf> (acessed on March 2015).

van Diepen, K., Boogaard, H., Supit, I., Lazar, C., Orlandi, S., Van der Goot, E., Schapendonk, A., 2004. Methodology of the MARS crop yield forecasting system. In: Lazar, C., Genovese, G. (Eds.), Agrometeorological Modelling, Processing and Analysis. EUR 21291 EN/2, p. 98 (accessed on March 2015) http://www.researchgate.net/publication/40795411_Methodology_of_the_MARS_crop_yield_forecasting_system._Vol._2_agrometeorological_data_collection_processing_and_analysis

van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance – a review. Field Crops Res. 143, 4–17.

Wu, B., Meng, J., Li, Q., Yan, N., Du, X., Zhang, M., 2014. Remote sensing-based global crop monitoring: experiences with China's CropWatch system. Int. J. Digital Earth 7 (2), 113–137.